DOCUMENT RESUME

ED 071 446 EM 010 696

| | |
|---|---|
| AUTHOR | Hegstrom, William J.; Phillips, William E. |
| TITLE | Comparability of Group Television and Teacher Administration of a Florida Reading Assessment Instrument. |
| INSTITUTION | Palm Beach County Board of Public Instruction, West Palm Beach, Fla. |
| PUB DATE | Jan 73 |
| NOTE | 9p.; Paper presented at the Annual Meeting of the Florida Educational Research Association (Tampa, Florida, January 1973) |
| EDRS PRICE | MF-$0.65 HC-$3.29 |
| DESCRIPTORS | Achievement Tests; *Elementary School Students; *Feasibility Studies; *Instructional Television; *Reading Tests; Standardized Tests |

ABSTRACT

A study was conducted to examine the feasibility of using instructional television as a means of standardizing countywide testing. Two reading assessment instruments were administered to second and fourth grade students through an oral instructional television reading of the items and through an oral reading by classroom teachers. Results showed that both second and fourth graders performed significantly better when the items were administered by teachers, although there were several explanations for these results which do not necessarily include that television may not be a better means of administering tests. (SH)

FILMED FROM BEST AVAILABLE COPY

ED 071446

# COMPARABILITY OF GROUP TELEVISION AND TEACHER ADMINISTRATION OF A FLORIDA READING ASSESSMENT INSTRUMENT

William J. Hegstrom
Supervisor, Research and Evaluation

William E. Phillips
Director of Learning Resources

School Board of Palm Beach County

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEH REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

A Paper Presented to the Annual Meeting of the Florida Educational Research Association, Tampa

January, 1973

EM 010 696

# COMPARABILITY OF GROUP TELEVISION AND TEACHER ADMINISTRATION OF A FLORIDA READING ASSESSMENT INSTRUMENT

William J. Hegstrom
Supervisor, Research and Evaluation

William E. Phillips
Director of Learning Resources

School Board of Palm Beach County

## Abstract

This pilot study investigated differences in the scores of two groups of 2nd graders and two groups of 4th graders who were tested using television administration vs teacher administration of the same reading assessment instrument. Null hypotheses concerning test scores for both grade levels were rejected in favor of the teacher-administration mode. Teacher-administrators appeared to prefer the teacher-administration mode while teacher-monitors in T.V. groups appeared to have a positive attitude toward T.V. testing.

This study was conducted to examine the feasibility of using the four channel I.T.V. network in Palm Beach County as a means of standardizing countywide testing. One objective was to test the hypothesis that scores obtained in group T.V. administration (T.V.A.) would not differ significantly from those obtained with teacher-administration (T.A.) of a Florida Reading Assessment instrument. The investigators believed that greater standardization and student effort would occur and resulting test scores might be raised if tests were administered over television.

## Related Studies

In a study using English-speaking Hawaiian students whose IQ's averaged 123, Fargo et al. (1967) found a slightly significant difference between "individually administered" and "T.V. administered" Peabody Picture Vocabulary Test scores. The results of the study favored "individual administration". Another study found no significant difference between "T.V." and

teacher administration" (Hopkins et al., 1967). Both studies reported implications pertaining to the standardization procedures for testing.

In comparing T.V. and audio testing Stallings (1972) found no significant difference in mean scores on the Modern Language Association French Listening Examination. But attitudes of students toward being tested by television was slightly unfavorable.

## Method

In May 1972, two reading assessment instruments, specially prepared from the Florida Reading Assessment sample items, were administered to second and fourth grade students in two groups of schools in Palm Beach County. Nine treatment and seven control group schools and teachers were selected from fifty-five (55) schools in a stratified-cluster sampling procedure. The experimental groups viewed a T.V. oral reading of the items in the second or fourth grade assessment instrument by a T.V. narrator (T.V.A.). Students responded to the T.V. narrator by checking a square under an appropriate picture in an answer booklet while their teachers monitored experimental classes. The control groups were administered the same second or fourth grade instruments by classroom teachers who read items to the students from identical scripts while students responded on identical answer booklets (T.A.). Hypotheses predicted no difference in test scores between experimental and control groups at each grade level. Two tailed t ratios were calculated to compare means at the .05 level of confidence.

In addition, an attitude questionnaire (Table I) was prepared and administered to the teachers of the control and experimental classes.

Twenty-four null hypotheses were proposed for the seven items and compiled total scores. Eight of them ($\mu_1 = \mu_2$) proposed no difference between experimental and control group scores in the totals and the seven items. The remaining sixteen ($P_1 = P_2 = P_3 = P_4 = P_5$) proposed that the five alternative choices were equally attractive in the totals and the 7 items for both groups.

TABLE I

TEACHER ATTITUDES TOWARD THE ADMINISTRATION OF STANDARDIZED EXAMS via I.T.V.

Hereinafter the administration of standardized exams over I.T.V. shall be called "T.V. testing". Please attempt to respond to each of the items below as you truly feel. Read every item first and then place a check mark (✓) in the column opposite each item which best describes your attitude.

(S.A. = Strongly Agree, A. = Agree, U. = Undecided, D. = Disagree, S.D. = Strongly Disagree)

| | 1 S.A. | 2 A. | 3 U. | 4 D. | 5 S.D. |
|---|---|---|---|---|---|
| 1. I approve of T.V. testing. | | | | | |
| 2. I prefer testing in a classroom with the administration done over a public address system. | | | | | |
| 3. I prefer testing in a classroom with administration by a teacher. | | | | | |
| 4. I prefer testing in large groups in a cafeteria. | | | | | |
| 5. I prefer testing in large groups in a library or an equally quiet location. | | | | | |
| 6. I believe T.V. testing involves too many problems such as equipment scheduling and breakdown. | | | | | |
| 7. I like the idea of T.V. testing using Spanish-speaking students. | | | | | |
| 8. List your suggestions for improving T.V. testing or other techniques. | | | | | |

Using the 1 - 5 scale a t ratio between experimental and control group means was calculated for each item and the total score (Ho: $X_E = X_C$, $\alpha = .05$). In calculating the total it was assumed that items two through six were the negative of responses to items one and seven and the scale was reversed accordingly.

The sixteen null hypotheses proposing that the five alternative choices (Likert scale) were equally attractive were tested using a chi-square goodness of fit statistic (Table II).

Results

A statistical comparison (t ratio) between T.V.A. and T.A. was calculated for each item and the total teacher attitude scale. The only significant difference in scores occurred on item #2 ($t = 2.7867$, $p < .05$). A goodness of fit $X^2$ test of the 1 - 5 scaled responses was then calculated using Yates correction for both T.V.A. and T.A. on all items and the totals. These results are shown in Table II on the following page.

Measures of cognitive responses by students on the specially prepared assessment instrument were calculated and are shown in Table III. The hypotheses predicting no difference in test scores between T.V.A. and T.A. for both grade levels were rejected.

TABLE II

$X^2$ Goodness of Fit Values of Each Item Over Each Group of Teachers on Attitudes Toward Exams via TV

| Item # | Group | $X^2$ | P | Question | Conclusion |
|---|---|---|---|---|---|
| 1 | Exp. | 5.7499 | > .10 | Approve T.V. testing | N.S. |
| | Cont. | 6.5156 | > .10 | " | N.S. |
| 2 | Exp. | 17.1718* | < .01 | Prefer P.A. System? | Opposes classroom administration using Public Address System. |
| | Cont. | 10.8364* | < .05 | " | |
| 3 | Exp. | 7.0832 | > .10 | Prefer classroom teachers administration? | N.S. |
| | Cont. | 35.9218* | < .01 | | Favors teacher administration |
| 4 | Exp. | 5.8332 | > .10 | Prefer large group in cafe.? | N.S. |
| | Cont. | 18.4218* | < .01 | | Opposes large group administration in cafeteria. |
| 5 | Exp. | 4.4165 | > .10 | Prefer large group in library? | N.S. |
| | Cont. | 9.0312 | < .10 | | Tends to oppose large group administration in a library. |
| 6 | Exp. | 5.8332 | > .10 | Feel that T.V. testing has too many problems? | N.S. |
| | Cont. | 12.7656* | < .02 | | Feel that T.V. testing involves too many problems. |
| 7 | Exp. | 5.8332 | > .10 | Prefer T.V. testing by Spanish narrator for Spanish student | N.S. |
| | Cont. | 7.4531 | > .10 | | N.S. |
| Total | Exp. | 20.9521* | < .01 | | Favors T.V. testing |
| | Cont. | 6.3132 | > .10 | | N.S. |

$N_{Exp.} = 15$
$N_{Cont.} = 16$

Yates correction for continuity was used in calculating $X^2$ values.

TABLE III

DISTRIBUTION OF SCORES FOR SECOND GRADE

| | Sample Size | Standard Deviation | Mean | t Ratio |
|---|---|---|---|---|
| T.A. | 142 | 2.31989 | 19.21 | 4.6529* |
| T.V.A. | 207 | 2.53285 | 17.99 | |

DISTRIBUTION OF SCORES FOR FOURTH GRADE

| | Sample Size | Standard Deviation | Mean | t Ratio |
|---|---|---|---|---|
| T.A. | 146 | 2.32538 | 17.36 | 4.2449* |
| T.V.A. | 223 | 2.75198 | 16.23 | |

*Significant at the .05 level

## Conclusion and Implications

In the absence of other controls it would appear that students in Palm Beach County do not do as well when a test is administered over television as they do when their teacher administers the test. Furthermore, there is some evidence that teachers of:

(1) both groups oppose classroom administration using a public address system;

(2) the control group favor classroom administration by the teacher;

(3) the control group oppose large group administration in a cafeteria;

(4) the control group feel that T.V. testing involves too many problems;

(5) the experimental group tend to indicate a positive attitude toward T.V. testing.

It may be that once teachers have used T.V. administration of tests, they are no longer quite so sure that they prefer self-administered testing.

The results of the study lead to a number of questions:

(1) Did the teachers unconciously cue the answers when reading the questions?

(2) Is it possible that teacher opposition to T.V. testing communicates to the students and effects his performance? (Rosenthal effect)

(3) Did T.V. time intervals, newness of the method of administration, etc., contribute to creating a student anxiety level which was too high?

(4) Does teacher attitude toward self-administration change as a result of exposure to T.V. testing?

(5) Is it possible that carefully controlled standardization of testing such as T.V.A. might lower national norms?

The use of an instrument which had been clinically validated but not field tested introduces some possibility of error. For that reason, it is suggested that a standardized test be administered to a sample of schools via I.T.V. on a similar follow-up study. The hypothesis is proposed that there would be no difference between the T.V. tested group and the teacher tested group.

To insure that no one misinterprets this study the reader is directed to the possibility that T.V. may be a better means of administering tests in spite of the results. More careful control over administration might result nationwide in a more valid set of norms. However, these results seem to indicate that standardized testing via T.V. should be limited to experimentation until more evidence has been gathered.

## BIBLIOGRAPHY

Fargo, G. A., Crowell, D. C., Noyes, M. H., Fuchigami, R. Y., Gordon, J. M. and Dunn-Rankin, P. Comparability of group television and individual administration of the Peabody Picture Vocabulary Test: Implications for screening. Journal of Educational Psychology, 1967, 58, 3, 137-140.

Hopkins, K. D., Lefever, D. W., and Hopkins, B. R. TV vs teacher administration of standardized tests: Comparability of scores. Journal of Educational Measurement, 1967, 4, 1, 35-40

Stallings, W. M. A comparison of television and audio presentations of the M.L.A. French Listening Examination. Journal of Educational Research, 1972, 65, 10, 472-74.